

# On the Complexity of Role Colouring Planar Graphs, Trees and Cographs

Christopher Purcell & M. Puck Rombach\*

February 2014

## Abstract

We prove several results about the complexity of the role colouring problem. A *role colouring* of a graph  $G$  is an assignment of colours to the vertices of  $G$  such that two vertices of the same colour have identical sets of colours in their neighbourhoods. We show that the problem of finding a role colouring with  $1 < k < n$  colours is NP-hard for planar graphs. We show that restricting the problem to trees yields a polynomially solvable case, as long as  $k$  is either constant or has a constant difference with  $n$ , the number of vertices in the tree. Finally, we prove that cographs are always  $k$ -role-colourable for  $1 < k \leq n$  and construct such a colouring in polynomial time.

## 1 Introduction

A *role colouring* of a graph  $G$  is an assignment of colours to the vertices of  $G$  such that two vertices of the same colour have identical sets of colours in their neighbourhoods. The concept arises from the study of social networks. Network science is an increasingly important application of graph theory and role colourings are a natural formulation of roles played by nodes in a real-world network [15, 16]. This structure was formalised by White and Reitz in terms of graph homomorphisms in [21], and developed extensively by Borgatti and Everett [2, 1, 6, 7]. A fast, applicable algorithm for finding role colourings is proposed in [11, 3]. A homomorphism  $h$  is said to be *locally surjective* if  $h$  is surjective when restricted to the neighbourhood set of any vertex. Locally surjective homomorphisms are equivalent to role colourings and they appear in the literature under many other names, *e.g.* role assignment [20], role equivalence [3], regular equivalence [2]. Throughout this paper we use the language of graph colourings and we refer to a role colouring using  $k$  colours as a  $k$ -role-colouring.

---

\*Department of Mathematics, University of California Los Angeles, 520 Portola Plaza, CA 90095, USA. [rombach@math.ucla.edu](mailto:rombach@math.ucla.edu), +1 (310) 206-9974.

We consider the computational problem associated with role colourings whose input is a graph  $G$  and whose output is a partition of the vertices of  $G$  into  $k$  non empty subsets satisfying the definition of a role colouring given above. We call this problem  $k$ -ROLE-COLOURABILITY, or  $k$ -ROLECOL for short. This problem differs from the more common COLOURABILITY problem in a few important ways. Firstly, every graph with no isolated vertex has a role colouring obtained by giving each vertex the same colour and every graph has a role colouring obtained by giving each vertex its own colour. Secondly a  $k$ -role-colouring does not usually imply the existence of a  $k + 1$ -role-colouring. This makes makes role colouring inherently different from the original graph colouring problem.

Finding role colourings of a given size is known to be NP-complete in general [14, 9]. For  $k \geq 3$ , the  $k$ -ROLECOL problem is NP-complete when restricted to chordal graphs [20]. However, 2-role-colouring can be solved in polynomial time for chordal graphs [18]. Not many other partial results on complexity of role colouring are known. In fact, interval graphs and trees are the only non trivial classes in which a polynomial solution is known to exist, and only for a constant number of colours.

The rest of this paper is organised as follows. In section 2, we prove that  $k$ -ROLECOL remains NP-complete even when restricted to planar graphs, a class that was suggested for examination in [20] and is one of the most extensively studied in the literature. In Section 3, we give an explicit algorithm that computes a  $k$ -role-colouring of a tree in polynomial time, as long as  $k$  is either constant or has a constant difference with  $n$ , the number of vertices in  $T$ . Finally, in Section 4, we show that every cograph (with at least  $k$  vertices) has a  $k$ -role-colouring, and hence that the decision version of the problem is solvable in polynomial time in this class. Our proof is constructive and gives an explicit algorithm to construct such a colouring.

## 2 Planar Graphs

In order to prove that  $k$ -ROLECOL is NP-complete when restricted to planar graphs, we introduce the SATISFIABILITY PROBLEM, defined below. A boolean formula  $\phi$  (in *conjunctive normal form*) is a set of clauses  $C_1, C_2, \dots$ , each of which is a set of variables  $x_1, x_2, \dots$ . The variables may take values TRUE or FALSE. For a given assignment of these values to the variables, a clause is said to be *satisfied* if at least one of its variables is assigned the value TRUE. A formula is satisfied if each of its clauses is satisfied. The SATISFIABILITY problem takes a boolean formula on  $n$  variables as its input and asks if there is an assignment of TRUE and FALSE to the variables that satisfies the formula. The general SATISFIABILITY problem was the first to be revealed to be NP-complete [4], and remains a central problem in theoretical computer science.

We will use a reduction from a certain restricted version of SATISFIABILITY. In order to describe this restricted problem, we define the following graph theoretic notion. The *formula graph*  $G_\phi$  of a given formula  $\phi$  is a bipartite graph whose vertices correspond to the clauses and variables of  $\phi$  with an edge be-

tween  $C$  and  $x$  if the variable  $x$  appears in the clause  $C$ . Let  $k$ -SATISFIABILITY be the SATISFIABILITY problem with the restriction that each clause contains at most  $k$  variables. The 3-SATISFIABILITY problem is NP-complete even when restricted to formulas with planar formula graphs [12]. In [19], Tovey showed that the problem is NP-complete under the restriction that each clause has two or three variables and each variable appears at most three times. We call the corresponding problem 3\*, 3\*-SATISFIABILITY. We now combine the restrictions imposed by Tovey and planarity to show that PLANAR 3\*, 3\*-SATISFIABILITY, which is 3\*, 3\*-SATISFIABILITY with planar formula graphs, is also NP-complete. We list a couple of planarity preserving operations that we will need throughout the coming proofs, in an easy lemma. See also [10].

**Lemma 1.** *If  $G'$  is a graph created from a planar graph  $G$  by any of the following operations, then  $G'$  is planar.*

- (a) *Adding a path  $x, z_1, \dots, z_k, y$  where  $x, y \in V(G)$ , and  $x, y$  share a face in some planar drawing of  $G$ , and  $z_1, \dots, z_k$  are new vertices.*
- (b) *Replacing a vertex  $x \in V(G)$  with  $d_G(x) = k$  by a cycle  $z_1, \dots, z_k$  with edges  $z_i, z_{i+1}$ ,  $1 \leq i \leq k-1$  and  $z_k, z_1$ , and edges  $z_i, y_i$ ,  $1 \leq i \leq k$ , where  $y_1, \dots, y_k$  are the neighbours in  $G$  of  $x$  appearing in clockwise order in a planar drawing of  $G$ .*
- (c) *Attaching a new planar subgraph  $H$  to  $G$ , such that  $V(G') = V(G) \cup V(H)$ ,  $E(G') = E(G) \cup E(H) \cup xz$ , where  $x \in V(G)$ ,  $z \in V(H)$ .*

*Proof.* (a) Replacing an edge by a multi-edge does not destroy planarity and replacing an edge  $xy$  by a path  $x, z_1, \dots, z_k, y$  clearly does not destroy planarity either.

(b) Cycles are planar and since the neighbours  $y_1, \dots, y_k$  and new vertices  $z_1, \dots, z_k$  are in the same order clockwise, the edges  $z_i, y_i$ ,  $1 \leq i \leq k$  do not cross each other or any new cycle-edges.

(c) We take the disjoint union of  $G$  and  $H$  and draw  $G$  such that  $x$  is on the outer face and  $H$  such that  $z$  is on the outer face. Adding an edge between two vertices on the same face does not destroy planarity. □

**Lemma 2.** *The PLANAR 3, 3-SATISFIABILITY problem is NP-complete.*

*Proof.* We follow the method [19] of reducing any 3-SATISFIABILITY problem to a 3\*, 3\*-SATISFIABILITY problem. The method is as follows. Suppose that variable  $x$  appears in  $k$  clauses. Create  $k$  new variables  $x_1, \dots, x_k$  and replace the  $i$ th occurrence of  $x$  with  $x_i$ , for  $1 \leq i \leq k$ . Append the clause  $\{x_i \vee \bar{x}_{i+1}\}$  for  $1 \leq i \leq k-1$  and  $\{x_k \vee \bar{x}_1\}$ . This new set of clauses forces the variables  $x_i$ ,  $1 \leq i \leq k$ , to be either all true or all false.

Let  $\phi$  be a PLANAR 3-SATISFIABILITY problem, with a given planar drawing of  $G_\phi$ . Suppose that variable  $x$  appears in  $k$  clauses. Create  $k$  new variables

$x_1, \dots, x_k$  and replace vertex  $x$  in  $G_\phi$  by the cycle on new vertices  $x_1, \dots, x_k$ . Replace the  $j_i$ th occurrence of  $x$  with  $x_i$ , for  $1 \leq i \leq k$ . Such that the clauses  $C_{j_1}, \dots, C_{j_k}$  appear in clockwise order in the neighbourhood of  $x$  in our given planar drawing of  $G_\phi$  (planarity-preserving operation (b) in lemma 1). Finally, replace each edge  $x_i x_{i+1}$  by a path  $x_i, C_{x_i}, x_{i+1}$ , for  $1 \leq i \leq k-1$ , and edge  $x_k x_1$  by path  $x_k, C_{x_k}, x_1$  (planarity-preserving operation (a) in lemma 1). The clause vertices  $C_{x_i}$  represent the new clauses  $\{x_i \vee \bar{x}_{i+1}\}$ . □

**Theorem 1.**  $k$ -ROLECOL,  $k \geq 2$ , is NP-hard for connected planar graphs.

*Proof.* Let  $\phi$  be a planar boolean formula on  $n$  variables  $x_1, x_2, \dots, x_n$  having  $m$  clauses  $C_1, C_2, \dots, C_m$ , such that each variable appears at most three times and each clause is of size 2 or 3. Let  $G_\phi$  be its formula graph. We will construct a related planar graph  $G'_\phi$  that has a  $k$ -role-colouring if and only if  $\phi$  is satisfiable. We split the proof into two cases.

First suppose  $k = 2$ . We construct  $G'_\phi$  from  $G_\phi$  as follows. To each clause vertex  $C_j$  we add a path  $a_j, b_j$  with edge  $b_j C_j$  (lemma 1 (c)). Since a variable appears in at most three clauses, at most one of  $x_i$  and  $\bar{x}_i$  appears twice. Furthermore, we can assume that a variable appears both positively and negatively. Subsequently, one of  $x_i$  and  $\bar{x}_i$  appears exactly once. If  $\bar{x}_i$  appears exactly once, in clause  $C_j$ , we replace the edge  $x_i C_j$  by a path  $x_i, \bar{x}_i, C_j$ . Otherwise, we have that  $x_i$  appears exactly once, in which case we relabel the node  $x_i$  to  $\bar{x}_i$ , and we replace the edge  $\bar{x}_i C_j$  by a path  $\bar{x}_i, x_i, C_j$  (lemma 1 (a)). Finally, we add a vertex  $y_i$  to each pair  $x_i$  and  $\bar{x}_i$  to form a triangle (lemma 1 (a)). See figure 1.

For each variable  $x_i$ , the graph  $G'_\phi$  will contain a copy of a triangle. With a slight abuse of notation for the sake of readability, we label these vertices  $x_i, \bar{x}_i, y_i$ . For each clause  $C_j$ ,  $G'_\phi$  contains a path on three vertices labelled  $a_j, b_j, C_j$ . Again this slight abuse of notation will aid the reader, and we refer to the vertices  $x_1, \bar{x}_1, \dots, x_n, \bar{x}_n$  as *literal vertices* and the vertices  $C_1, \dots, C_m$  as *clause vertices*. We add an edge between a literal vertex  $x$  and a clause vertex  $C$  if the literal  $x$  appears in the clause  $C$ .

Suppose that  $G'_\phi$  has a 2-role-colouring. Without loss of generality, the vertex  $a_1$  is red. It is easy to see that  $b_1$  cannot also be red, for otherwise red vertices could only have red neighbours, and every vertex would then be red which would be a contradiction. Suppose  $C_1$  is red. Then, since  $b_1$  is blue and has only red neighbours, the neighbourhood of every blue vertex must be red. Let  $x$  be a neighbour of  $C_1$  other than  $b_1$ . Since  $x$  is a neighbour of a red vertex it must be coloured blue. But  $x$  is contained in a triangle, the other two vertices of which must be coloured red. This contradiction shows that  $C_1$  must be coloured blue, and we can deduce that each of the paths  $a_j, b_j, C_j$  must be coloured red, blue, blue from the fact that the vertex  $a_j$  has degree 1 and since a blue vertex must have at least one neighbour of each colour,  $a_j$  must be coloured red.

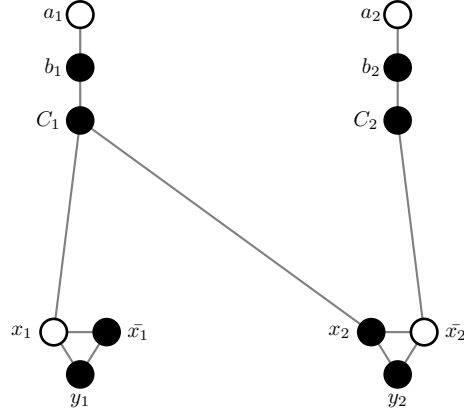


Figure 1: A graph  $G_\phi$  representing the boolean formula  $\phi = \{x_1 \cup x_2\} \cap \{\bar{x}_2\}$  with a 2-role colouring corresponding to a satisfying assignment where  $x_1$  and  $\bar{x}_2$  are true.

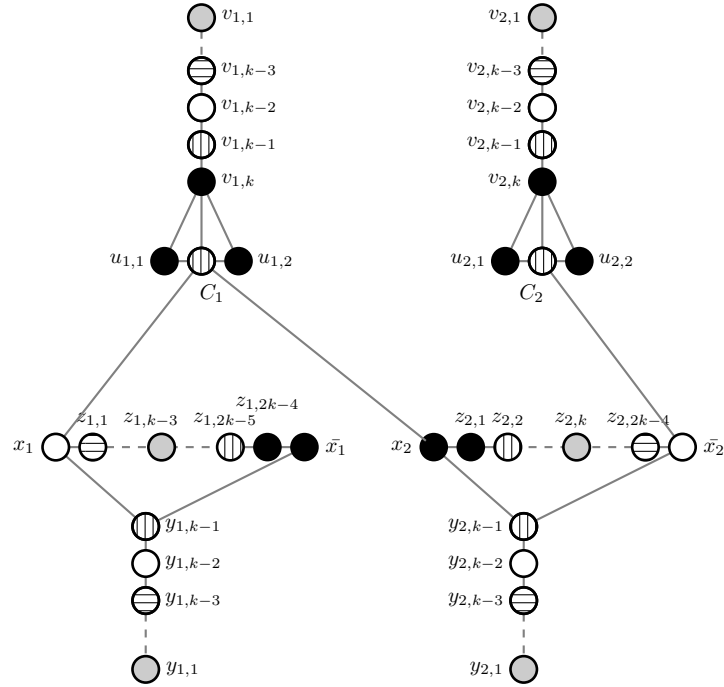


Figure 2: A graph  $G_\phi$  representing the boolean formula  $\phi = \{x_1 \cup x_2\} \cap \{\bar{x}_2\}$  with a  $k$ -role colouring corresponding to a satisfying assignment where  $x_1$  and  $\bar{x}_2$  are true.

Now consider the triangle  $x_1, \bar{x}_1, y_1$ . They cannot all be coloured blue because blue vertices must have red neighbours. No two of them can be red, because red vertices cannot have red neighbours. Therefore exactly one of  $x_1, \bar{x}_1, y_1$  is coloured red. Observe that each clause vertex must have a red neighbour amongst the literal vertices in its neighbourhood. We may therefore construct a truth assignment by assigning TRUE to the variable  $x_i$  if and only if the literal vertex  $x_i$  is red.

In order to prove the result for  $k > 2$  we use a slightly different construction, and introduce the following notation. An induced path  $\{v_1, v_2, \dots, v_k\}$  in a graph  $G$  is said to be *dangling* if  $v_1$  is of degree 1 in  $G$  and  $v_2, \dots, v_{k-1}$  are of degree 2 in  $G$ .

**Lemma 3.** *If a connected graph  $G$  has a dangling path  $P$  of length at most  $k$ , then in any  $k$ -role-colouring of  $G$  no two vertices of  $P$  can have the same colour*

*Proof.* Let  $\{v_1, v_2, \dots, v_k\}$  be a dangling path in  $P$  such that  $v_1$  has degree 1. We have observed that the closed neighbourhood of a vertex cannot be monochromatic. Without loss of generality,  $v_1$  has colour 1 and  $v_2$  has colour 2. Suppose that  $v_i$  has colour  $i$  for  $i < j$ . Clearly  $v_j$  cannot have colour  $c < j - 2$ . Suppose  $v_j$  has colour  $j - 2$  or  $j - 1$ . Then no vertex of colour at least  $j$  can have a neighbour of colour at most  $j - 1$ . This contradicts the connectedness of  $G$ . Therefore without loss of generality,  $v_j$  has colour  $j$ .  $\square$

We now give a description of  $G'_\phi$  in the case that  $k$  is at least 3. Let  $\phi$  be a PLANAR 3,3-SATISFIABILITY problem on  $n$  variables  $x_1, x_2, \dots, x_n$  having  $m$  clauses  $C_1, C_2, \dots, C_m$ , and let  $G_\phi$  be its formula graph. We construct a related planar graph  $G'_\phi$  that has a  $k$ -role-colouring if and only if  $\phi$  is satisfiable.

We construct  $G'_\phi$  from  $G_\phi$  as follows. To each clause vertex  $C_j$  we add a dangling path  $v_{j,1}, \dots, v_{j,k}$ , with edge  $v_{j,k}, C_j$  (lemma 1 (c)). To each pair  $v_{j,k}, C_j$  we add two vertices  $u_{j,1}, u_{j,2}$  that both form a triangle with  $v_{j,k}, C_j$  (lemma 1 (a)). As before, one of  $x_i$  and  $\bar{x}_i$  appears exactly once. If  $\bar{x}_i$  appears exactly once, in clause  $C_j$ , we replace the edge  $x_i C_j$  by a path  $x_i, z_j, 1, \dots, z_j, 2k - 4, \bar{x}_i, C_j$ . Otherwise we have that  $x_i$  appears exactly once, in which case we relabel the node  $x_i$  to  $\bar{x}_i$ , and replace the edge  $\bar{x}_i C_j$  by a path  $\bar{x}_i, z_j, 1, \dots, z_j, 2k - 4, x_i, C_j$  (lemma 1 (a)). We add a vertex  $y_{i,k-1}$  and attach it to  $x_i$  and  $\bar{x}_i$  (lemma 1 (a)). Finally, we add a dangling path  $y_{i,1}, \dots, y_{i,k-1}$  to the graph through the edge  $y_{i,k-1}, y_{i,k}$  (lemma 1 (c)). See figure 2.

As in the first part of the proof, suppose  $G'_\phi$  has a  $k$ -role-colouring. Observe that the path formed by  $v_{i,1}, \dots, v_{i,k}$  is dangling, and must therefore be coloured with  $k$  distinct colours. Without loss of generality  $v_{1,1}$  has colour 1, and indeed  $v_{1,j}$  has colour  $j$ . Therefore the neighbours of any vertex of colour  $j$  must have colour  $j - 1$  or  $j + 1$  for  $1 < j < k$ . Consider the vertices  $u_{1,1}, u_{1,2}$ . If either of them has colour  $k - 1$ , then  $C_1$  must have colour  $k - 2$ . But  $C_1$  is adjacent to  $v_{1,k}$  which has colour  $k$  which leads to a contradiction. So  $u_{1,1}$  and  $u_{1,2}$  must have colour  $k$  and therefore  $C_1$  has colour  $k - 1$ .

Since the role-graph of this colouring of  $G'_\phi$  is a simple path on the colours  $1, \dots, k$  with  $k$  having a self loop, all vertices of degree 1 must have colour 1. So

for each clause  $C_i$ , we have that  $v_{i,1}$  has colour 1,  $v_{i,2}$  has colour 2, and so on. This implies that every subgraph induced on vertices  $C_i, v_{i,1}, \dots, v_{i,k}, u_{i,1}, u_{i,2}$  has the same colouring as described above for  $i = 1$ .

For each variable  $x_i$ , we have that  $y_{i,1}$  has colour 1,  $y_{i,2}$  has colour 2, ...,  $y_{i,k-1}$  has colour  $k-1$ . This implies that the vertices  $x_i$  and  $\bar{x}_i$  must have colours  $k-2$  and  $k$ , or vice versa. If  $x_i$  receives colour  $k-2$  and  $\bar{x}_i$  receives colour  $k$ , then the vertices  $z_{i,1}$  through  $z_{i,k-3}$  through  $z_{i,2k-4}$  must receive colours  $k-3$  through 1 through  $k$ . Alternatively, if  $x_i$  receives colour  $k$  and  $\bar{x}_i$  receives colour  $k-2$ , then the vertices  $z_{i,1}$  through  $z_{i,k-3}$  through  $z_{i,2k-4}$  must receive colours  $k$  through 1 through  $k-3$ .

Now we construct a satisfying assignment for  $\phi$  from this colouring. If the vertex representing  $x_i$  has colour  $k-2$  we assign the variable  $x_i$  the value TRUE. If the vertex representing  $\bar{x}_i$  has colour  $k-2$  we assign the variable  $\bar{x}_i$  TRUE.

Since each vertex representing a clause must have a neighbour of colour  $k-2$ , each clause now has a variable or its negation that has been assigned TRUE, and therefore  $\phi$  is satisfied. □

### 3 Trees

Let  $P_m$  denote a path of length  $m$ , *i.e.* a graph on vertex set  $1, \dots, m+1$  with edges  $(i, i+1)$  for  $1 \leq i \leq m$ . Let  $T$  denote a tree, *i.e.* a connected graph on  $n$  vertices and  $n-1$  edges with no cycles.

For a valid role colouring of the vertices of a graph  $G$  using  $k$  colours, the *role graph*  $G^R$  is defined in an obvious way.  $G^R$  has  $k$  vertices, each one corresponding to a colour used on  $V(G)$ . Vertices  $i$  and  $j$  in  $V(G^R)$  have an edge in  $G^R$  if and only if vertices of colour  $i$  are always connected to a vertex of colour  $j$  in  $G$ . The graph  $G^R$  may have self-loops. It is easy to see that  $\Delta(G^R) \leq \Delta(G)$ ,  $\delta(G^R) \leq \delta(G)$ , where  $\delta(G)$  and  $\Delta(G)$  denote the minimum and maximum degree in  $G$ , respectively, over all vertices. If  $G$  is connected, then  $G^R$  must be connected. (The converse is not true.) This holds because a path in  $G$  must correspond to a walk in  $G^R$ .

**Lemma 4.** *A path  $P_{n-1}$  can be role coloured using  $k$  colours if and only if  $n = k + s(k-1)$  or  $n = 2k + s(2k-1)$ , where  $s$  is a positive integer, and PATH  $k$ -ROLE COLOURABILITY is in  $P$ .*

*Proof.* By the properties of  $G^R$ , we see that  $P_{n-1}^R$  must be a path. It may have one self-loop on a leaf vertex. The path from vertex 1 to vertex  $n$  on  $P_{n-1}$  corresponds to a walk on  $P_{n-1}^R$  that must start and end at a vertex of degree one. If  $P_{n-1}^R$  contains no self-loops then such walks can be of length  $k, k + (k-1), k + 2(k-1), \dots$ . If  $P_{n-1}^R$  contains one self-loop then such walks can be of length  $2k, 2k + (2k-1), 2k + 2(2k-1), \dots$ .

Checking whether either of these equalities holds is clearly in  $P$ , and then colouring the path is in  $P$ , because for a given  $P_{n-1}^R$ , there are at most two ways of colouring  $P_{n-1}$ . □

**Lemma 5.** *In a role colouring of a tree  $T$ ,  $T^R$  must be a tree with at most one self-loop.*

*Proof.* Let  $T^R$  be a role graph with a self-loop on vertex  $w \in V(T^R)$ . Consider an edge  $(v_1, v_2)$  in  $T$  connecting two vertices of colour  $w$ . If we cut this edge, we have two components that must each contain a subgraph isomorphic to  $T^R$  without the self-loop on  $w$ . However, if  $T^R$  contains another self-loop, this process must repeat infinitely many times.  $\square$

**Lemma 6.** *For trees,  $k$ -ROLE COLOURABILITY is in  $P$ , if  $k$  is constant or if  $n - k$  is constant.*

*Proof.* It is shown in [8] that, for a tree  $T$ , and a known role graph  $T^R$  without self-loops, checking role colourability can be done in polynomial time<sup>1</sup>. By Cayley's formula, there are  $k^{k-2}$  labelled trees on  $k$  vertices, and  $(k+1)k^{k-2}$  trees with one or no self-loops. Therefore, one can check colourability for all possible role graphs, of which there are a constant number. Let  $T^R$  be a role graph with a self-loop on vertex  $v \in V(T^R)$ . Consider  $T_*^R$ , which is constructed as follows. Take two copies of  $T^R$  without the self-loop, one with vertex set  $1, \dots, w, \dots, k$  and a copy with vertex set  $1', \dots, w', \dots, k'$ . Let  $T_*^R$  be the union of these two trees with an edge added between  $w$  and  $w'$ . A valid  $2k$ -role colouring of  $T$  according to  $T_*^R$  now corresponds to a valid  $k$ -role colouring according to  $T^R$  by merging the colour classes 1 and  $1'$ , 2 and  $2'$ , etc.

**Claim 1.** *Let  $k' = n - k$ . Suppose  $T$  has a valid  $k$ -role colouring. Let  $v_1, \dots, v_t$  be a path where  $v_1$  and  $v_t$  are vertices of the same colour. Then this path contains vertices of no more than  $\lceil t/2 \rceil$  colours. Additionally, if  $v_1$  and  $v_t$  are removed from  $T$ , we are left with three components,  $T_a, T_b, T_c$ , where  $T_b$  contains the path  $v_1, \dots, v_t$ . Then  $T_a$  and  $T_c$  must contain the same colour set.*

*Proof.* Vertices  $v_2$  and  $v_{t-1}$  must have the same colour. If not, then the path would correspond to a cycle in  $T^R$ , which is a contradiction. This argument can be repeated for the path  $v_2, \dots, v_{t-1}$ . The neighbours of  $v_1$  and  $v_t$  that are not on the path  $v_1, \dots, v_t$  must have the same colour sets.

Without loss of generality, suppose there are vertices in  $T_a$  of a colour that does not appear in  $T_c$ . Let  $v_a$  be such a vertex that is the closest to  $v_1$ . The second vertex on the path between  $v_a$  and  $v_1$  is of a colour that appears in  $T_c$ , but is adjacent to a vertex of a colour that does not. This is a contradiction.  $\square$

**Claim 2.** *If  $T$  is  $k$ -role coloured with  $k' = n - k$  a constant, then a vertex which cuts  $T$  into more than one component of size  $> 2k' + 1$  must have a unique colour.*

*Proof.* Follows directly from claim 1.  $\square$

---

<sup>1</sup>The algorithm given is only geared towards deciding role colourability, but it is easily transformed into an algorithm that finds an explicit colouring in polynomial time.



We define a *gadget* as follows. A gadget is a maximal subtree of  $T$  of size at most  $2k' + 1$  such that the complement of the gadget in  $T$  is connected. The vertices that are adjacent to a gadget but are not in a gadget themselves are called *hubs*. Gadgets are clearly non-overlapping, and a hub may have many gadgets connected to it. These definitions are illustrated in figure 3.

**Claim 3.** *Repeating colours can only appear within gadgets and two vertices of the same colour are either in the same gadget or in gadgets adjacent to the same hub.*

*Proof.* This follows from claims 1 and 2. □

Colour each hub with a unique colour. For every hub, go through each gadget and record all possible colourings of the subgraphs induced on the gadget and the hub, and corresponding gadget role graph (which includes a node for the unique colour of the hub). For any gadget this can be done in constant time. Additionally, keep track of all combinations of multiple gadgets that can be role coloured using the same gadget role graph. This can all be done by brute force in polynomial time, as there are  $O(n)$  gadgets,  $O(1)$  different gadget role graphs. Now, record a list of all the possible numbers of duplicate-coloured vertices within the hub and its gadgets. Suppose the gadgets are coloured one by one, such that those with duplicate colours are coloured first. There are  $O(n - k) = O(1)$  such gadgets with a constant number of possible role colourings each. After these gadgets have been role coloured the other gadgets must be rainbow coloured. Therefore, recording all possible role colourings of the hub and its gadgets that yield no more than  $k'$  duplicate-coloured vertices takes  $O(n^{k'})$  time.

Once all possible colourings for the individual hubs and their gadgets have been recorded, consider all combinations of different hub and gadget colourings. Each hub and its gadgets under one of these colourings contains  $\leq k'$  duplicate-coloured vertices, and we have found a successful colouring if there is a set of hub and colouring combinations that add up to  $k'$  duplicate-coloured vertices. We need  $O(1)$  hubs in such a combination so, again, a brute force search of all combinations is sufficient. If such a combination does not exist, then a  $k$ -role colouring does not exist. Otherwise, colour the relevant hubs and their gadgets successfully, and rainbow-colour the remaining uncoloured vertices in  $T$ . This results in a valid  $k$ -role colouring of  $T$ . □

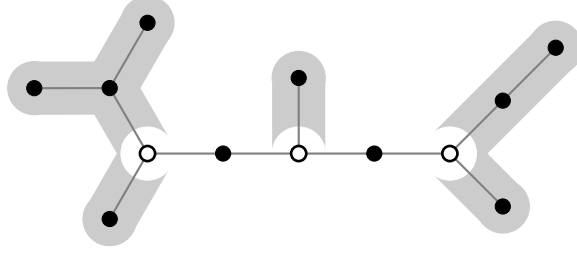


Figure 3: Hubs are the vertices that separate gadgets from the rest of the tree. Here  $k' = 1$ , hubs are white-filled and gadgets are shaded in grey.

## 4 Cographs

*Cographs* are exactly the  $P_4$  free graphs [17]. The join of two graphs  $G_1$  and  $G_2$  is the graph  $G_3 = G_1 + G_2$  such that  $V(G_3) = V(G_1) \cup V(G_2)$  and  $E(G_3) = E(G_1) \cup E(G_2) \cup \{(i, j) | i \in V(G_1), j \in V(G_2)\}$ . The disjoint union of two graphs  $G_1$  and  $G_2$  is the graph  $G_3 = G_1 \cup G_2$  such that  $V(G_3) = V(G_1) \cup V(G_2)$  and  $E(G_3) = E(G_1) \cup E(G_2)$ . Cographs can be constructed recursively from  $K_1$  by disjoint union and join. They are the smallest class of graphs closed under disjoint union and join. Every cograph  $G$  has an associated (binary, not necessarily unique) *cotree*, whose leaves correspond to the vertices of  $G$ , and the non-leaves are labelled “0” and “1” to denote disjoint unions and joins, respectively. The cotree describes how  $G$  is formed from instances of  $K_1$  by successive joins and disjoint unions. Given a cograph, its cotree can be found in linear time [5, 13].

**Theorem 2.** *All cographs with  $\geq 2$  vertices are 2-role-colourable and 2-ROLECOL for cographs is in P.*

*Proof.* Suppose  $G$  is a cograph. If  $G$  is not connected, then we can mono-colour each component either red or blue, such that both red and blue are used and such that, if  $G$  contains both isolated vertices and components of size  $\geq 2$ , these two types of components receive different colours.

Therefore, suppose  $G$  is connected. A connected cograph  $G$  with  $|V(G)| = 2$  is isomorphic to  $K_2$ , and can both be 2-role-coloured in the obvious way. Suppose that all connected cographs  $G'$  with  $|V(G')| < k$  can be 2-role-coloured. Suppose  $|V(G)| = k > 2$  and the last step in the construction of  $G$  was a join of graphs  $G_1$  and  $G_2$  (which it must be if  $G$  is connected). We consider three separate cases.

- (i) If  $|V(G_1)|, |V(G_2)| > 1$ , then we can 2-role-colour the vertices of  $G_1$  (red and blue) and  $G_2$  (red and blue). This extends to a valid 2-role-colouring of  $G$ , because all vertices have red and blue neighbours.
- (ii) If  $|V(G_1)| = 1, |V(G_2)| > 1$  (without loss of generality) and  $G_2$  is 1-role-colourable, then colour the vertex of  $G_1$  red and the vertices of  $G_2$  blue.

This extends to a valid 2-role-colouring of  $G$ , because the red vertex has only blue neighbours and the blue vertices have only a red neighbour or red and blue neighbours, depending on whether  $G_2$  is empty or not.

- (iii) If  $|V(G_1)| = 1$ ,  $|V(G_2)| > 1$  (without loss of generality) and  $G_2$  is not 1-role-colourable, then colour the vertex of  $G_1$  red. The graph  $G_2$  is not 1-role-colourable, which means that it is disconnected with isolated vertices and components with  $\geq 2$  vertices. Colour the isolated vertices of  $G_2$  blue. For each component of  $G_2$  with  $\geq 2$  vertices, colour one vertex blue and the others red. This extends to a valid 2-role colouring of  $G$ , because all blue vertices have only red neighbours and all red vertices have both red and blue neighbours.

Therefore, we can always find a valid 2-role-colouring for  $G$ . It is easy to see that this method can be executed in polynomial time. We can find a cotree in polynomial time, which gives us a  $G_1$  and  $G_2$ . Then we find the connected components of  $G_1$  and  $G_2$ , which can also be done in polynomial time (by a series of at most  $n$  breadth first searches).  $\square$

**Theorem 3.** *All cographs with  $\geq k$  vertices are  $k$ -role-colourable and  $k$ -ROLECOL for cographs is in  $P$ , where  $k > 2$ .*

*Proof.* We know that all cographs with  $\geq 2$  vertices are 2-role-colourable, so suppose that all cographs are  $k'$ -role-colourable for all  $2 \leq k' < k$ . Suppose  $G$  is a cograph with  $|V(G)| = n \geq k$  and the last step in a construction of  $G$  was either a join or a disjoint union of  $G_1$  and  $G_2$ , with  $|V(G_1)| = n_1$  and  $|V(G_2)| = n_2$ . Pick  $k_1$  and  $k_2$  such that  $k_1 + k_2 = k$ ,  $k_1 \leq n_1$ ,  $k_2 \leq n_2$  and  $k_i = 1$  only if  $n_i = 1$ , for  $i = 1, 2$ . Now,  $G_i$  is  $k_i$ -role-colourable by our inductive assumption, for  $i = 1, 2$ . Note that  $k_i$  is only equal to 1 if  $n_i$  is 1, and  $K_1$  is always 1-role-colourable. So, we can colour  $G_1$  using  $k_1$  colours and  $G_2$  using  $k_2$  different colours. This extends to a valid  $k$ -role-colouring of  $G$  regardless of whether  $G = G_1 \cup G_2$  or  $G = G_1 + G_2$ .  $\square$

## 5 Acknowledgements

Puck Rombach is supported by AFOSR MURI grant FA9550-10-1-0569 and ONR grant N000141210040. Part of this work was undertaken while Puck Rombach was attending the semester program “Network Science and Graph Algorithms” at the Institute for Computational and Experimental Research in Mathematics (ICERM) at Brown University.

## References

- [1] Stephen P Borgatti and Martin G Everett. Notions of position in social network analysis. *Sociological Methodology*, 22(1):1–35, 1992.

- [2] Stephen P Borgatti and Martin G Everett. Two algorithms for computing regular equivalence. *Social Networks*, 15(4):361–376, 1993.
- [3] Ronald S Burt. Detecting role equivalence. *Social Networks*, 12(1):83–97, 1990.
- [4] Stephen A Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pages 151–158. ACM, 1971.
- [5] Alain Cournier and Michel Habib. A new linear algorithm for modular decomposition. In *Trees in Algebra and Programming—CAAP’94*, pages 68–84. Springer, 1994.
- [6] Martin G Everett and Stephen P Borgatti. Regular equivalence: General theory. *Journal of Mathematical Sociology*, 19(1):29–52, 1994.
- [7] Martin G Everett and Steve Borgatti. Role colouring a graph. *Mathematical Social Sciences*, 21(2):183–188, 1991.
- [8] Jiří Fiala and Daniël Paulusma. Comparing universal covers in polynomial time. In *Computer Science—Theory and Applications*, pages 158–167. Springer, 2008.
- [9] Jiří Fiala and Daniël Paulusma. A complete complexity classification of the role assignment problem. *Theoretical Computer Science*, 349:67–81, 2005.
- [10] Jonathan L Gross and Jay Yellen. *Graph Theory and its Applications*. CRC press, 2005.
- [11] Hans J Hummell and Wolfgang Sodeur. Strukturbeschreibung von positionen in sozialen beziehungsnetzen. *Methoden der Netzwerkanalyse*, pages 177–202, 1987.
- [12] David Lichtenstein. Planar formulae and their uses. *SIAM Journal on Computing*, 11:329–343, 1982.
- [13] Ross M McConnell and Jeremy P Spinrad. Linear-time modular decomposition and efficient transitive orientation of comparability graphs. In *Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 536–545. Society for Industrial and Applied Mathematics, 1994.
- [14] Fred S Roberts and Li Sheng. How hard is it to determine if a graph has a 2-role assignment? *Networks*, 37:67–73, 2001.
- [15] Ryan A Rossi and Nesreen K Ahmed. Role discovery in networks. *arXiv preprint arXiv:1405.7134*, 2014.
- [16] Lee Douglas Sailer. Structural equivalence: Meaning and definition, computation and application. *Social Networks*, 1(1):73–90, 1979.

- [17] Dieter Seinsche. On a property of the class of  $n$ -colorable graphs. *Journal of Combinatorial Theory, Series B*, 16(2):191–193, 1974.
- [18] Li Sheng. 2-role assignments on triangulated graphs. *Theoretical Computer Science*, 304:201–214, 2003.
- [19] Craig A. Tovey. A simplified NP-complete satisfiability problem. *Discrete Applied Mathematics*, 8(1):85–89, 1984.
- [20] Pim van’t Hof, Daniël Paulusma, and Johan M M van Rooij. Computing role assignments of chordal graphs. *Theoretical Computer Science*, 411(40):3601–3613, 2010.
- [21] Douglas White and Karl Reitz. Graph and semigroup homomorphisms on networks of relations. *Social Networks*, 5:193–235, 1983.